1965

Serial correlation in pseudo-random numbers

Kenneth J. Lamport
Lehigh University

SERIAL CORRELATION IN PSEUDO-RANDOM NUMBERS

by

Kenneth J. Lamport

A THESIS

Presented to the Graduate Faculty

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Industrial Engineering

Lehigh University
1965

CERTIFICATE OF APPROVAL

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

Sept. 20, 1965
(date)

John M. Carroll
Professor in Charge

Head of the Department

ACKNOWLEDGEMENTS

My sincere thanks to John M. Carroll, my Thesis Advisor, for his guidance and enthusiasm and to Sutton Monro for his generous help in the statistical design.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The statistical properties of pseudo-random numbers generated by the mixed congruential technique are the subject of much deliberation. One characteristic discussed by several authors is the serial correlation within the numbers of a generated sequence. This property is discussed, however, on a theoretical basis only, and its effects upon a computer simulation are not recorded in the literature. It is the thesis here that serial correlation does exist in pseudo-random generated sequences and that this characteristic can bias the results of a simulation.

The approach developed to isolate mixed generators exhibiting serial correlation is taken from the classical techniques of time series analysis. A correlation statistic is computed and its statistical significance is determined. The frequency characteristics of each generator tested are also ascertained and specific generators are then chosen for sample simulations. The simulation is a Monte Carlo integration to determine the normalized volumes of n-dimensional spheres. Results of the simulations are then tested for significance.

Upon evaluation of the results of the simulations and the tests, it is found that serial correlation of both low and high order does exist in pseudo-random sequences and that it does bias, although not with statistical significance, the results of the simulations.

PREFACE

The technique of simulation should interest any Industrial Engineer who deals with problems for which mathematical models and probability distributions aid in the conventional solution. Basic to any analysis by simulation is a set of random or pseudo-random numbers. Generation of these numbers as they are needed is found to be the most convenient way of obtaining them in computer simulations. The purpose of this thesis is to study one property of these computer generated pseudo-random numbers, serial correlation, and to relate this property to the results of a simulation analysis.

CHAPTER 1: INTRODUCTION

1.1 The Random Number Generator

The most popular techniques for the computer generation of pseudo-random numbers are the multiplicative and mixed congruential methods. This is due to the speed of generation, the good statistical properties, and the long periods obtained from the congruential generators. The mixed generator will be the object of study in this thesis.

The mixed congruential pseudo-random number generator, which will hereafter be the generator referred to, is of the form

$$x_{n+1} \equiv ax_n + c \pmod{m} \qquad (1)$$

where x_n is the nth pseudo-random number generated, x_{n+1} is the next number of the sequence, a is a constant of the form 2^s+1, c is any odd number, and m is expressed as 2^L for a binary computer of word length L. The congruence sign implies the following: x_n is multiplied by a and added to c; the result is divided by m; x_{n+1} is the remainder after the division has been performed. Random numbers between zero and one are the most easily adaptable for transformation to the appropriate distribution of the simulation. Because of this, only pseudo-random sequences between zero and one, excluding one due to the round-off procedure of the generator program, will be considered.

Most important in the selection of pseudo-random number generators are the length of period and the statistical behavior of the generated sequence. There is good agreement in the literature as to the parameters which constitute a generator of maximal period (1,2,3,4,5,6). These parameters are quite clearly defined by Hull and Dobell (1, p. 233) who state that the maximum period, m, can be obtained provided that:

a. c and m are relatively prime
b. $a \equiv 1 \pmod{p}$ if p is a prime factor of m
c. $a \equiv 1 \pmod{4}$ if 4 is a factor of m.

For the GE 225, which is a binary machine, the above requirements demand only that c be odd and that a be taken from a = 4k 1 where k = 1,2,3..... Any starting value, x_0, may be chosen for the generator.

Statistically, it is desired that the pseudo-random sequence exhibit perfectly random properties. Each number in the sequence should show neither too much or too little relationship with every other number; but often a correlation will occurr between one number and a following number a set distance or lag away. The generated sequence should be uniformly distributed over the interval (0,1); but a non-uniform frequency of occurrence is quite common among pseudo-random numbers. Periodic properties are known to appear in pseudo-random sequences, and cycling will occurr if the length of the period is not sufficient. Also possible are sub periods and harmonic properties sometimes of predictable length (4, p. 611). The statistical acceptability of the pseudo-random sequences, therefore, is the subject of much deliberation. But before deliberation on the mixed congruential pseudo-random number generator can begin, the criteria of judgement must be established and agreed upon. Two such criteria often used are the frequency test and the serial correlation test. Many other criteria exist but consideration of them will be omitted.

Hull and Dobell (2, p. 36), for example, report unacceptable frequency and serial correlation characteristics for mixed generators with $x_0 = 0$, $c = 1$, $m = 2^{35}$, and $s = 12$. Very small multipliers, $a = 1, 5, 9, 17$, are said to show unsatisfactory serial correlation although all except $a = 1$ pass the frequency criterion. Some improvement is noted for more complicated values of c (2, p. 39), but changing x_0 and c is reported to have very little effect upon the acceptability of the generator.

Other reports include that of Rotenberg (5, p. 75) who tested the mixed generator with $s = 7$, $c = 1$, and $m = 2^{35}$ for 4096 numbers and found it acceptable for frequency as well as for other criteria. Peach (4, pp. 612-616) used $s = 11$, $c = 211,527,139$, $m = 2^{28}$, and $x_0 = 0$ for over 250,000 numbers and found undesirable patterns although satisfactory frequency results. He concluded that his sequence of numbers contained both patterns and periodicities which acted as constriants upon the variability of the sequence.

It must be remembered that the results of the above experiments, as well as of those to be described here, depend very much upon the tests which are used to determine acceptability.

1.2 Serial Correlation

Theoretical Approach

An indication of the distribution of the values in a sequence of pseudo-random numbers is given by the serial correlation of the sequence (6, p. 79). Considerable work has been done for the theoretical determination of serial correlation for pseudo-random sequences generated by the mixed congruential technique. Coveyou (7, p. 73) has shown that the correlation coefficient, p, between x_n and x_{n+1} for a sequence of full period can be expressed by

$$p(x_n, x_{n+1}) = \frac{1 - 6(c/m)(1 - c/m)}{a} \tag{2}$$

The assumptions underlying this development are considerably weakened for small a and c.

It is quite possible that the correlation between x_n and x_{n+k}, where k is any number greater than or equal to one and represents the lag of the correlation, is also important. For such considerations Coveyou has shown that formula (2) can be used if one may write

$$x_{n+k} \equiv a_k x_n + c_k \quad (\text{mod } m) \tag{3}$$

with $\quad a_k \equiv a^k \ (\text{mod } m).$ (4)

and $\quad c_k \equiv ((a^k - 1)/(a-1))c(\text{mod } m).$ (5)

Greenberger (6, p. 80) has derived a correction term to be added to Coveyou's approximation which is given by

$$T = (12/m)(S/a^2 - a/4) \tag{6}$$

Where S is a function of parameters a, c, and m of the generator formula.

Tocher (6, p. 80) states that Greenberger has developed methods for determining S for any given generator and gives as the results of computations of serial correlation of lag one:

$p=\frac{1}{4}$ for $m=2^{35}$, $a=2^{34}+1$, and $c=1$

$p=2^{-18}$ for $m=2^{35}$, $a=2^{18}+1$ and $c=1$.

The serial correlation coefficient of $\frac{1}{4}$ indicates some degree of correlation while the second coefficient, 2^{-18}, shows exceptional independence or non correlation of the generated numbers.

There are several obvious limitations in the above approach to the study of serial correlation. First, the formulas hold true only for full period m whereas partial samplings from a sequence are normally used in a computer simulation. Thus, the serial correlation of partial samplings must still be investigated by conventional techniques (8, p. 166). Second, no statistical significance can be attached to serial correlation coefficients obtained in the above manner. Lastly, the approach discussed above has been carried out on a theoretical basis only. No mention has been made in the literature of the results of a simulation carried out with serial correlation inherent in the generated sequence. It is the thesis here that serial correlation could have a significant effect upon such results.

Time Series Approach

Hoel (9, p. 342), in reference to non-parametric time series analysis, states that:

> "If observations have been ordered with respect to time and time is irrelevant, no correlation would be expected to exist, for example, between successive pairs of values of the sequence. However, if there is a cyclical movement in the sequence, neighboring pairs of values will tend to be high or low together and thus produce a value of the correlation coefficient that differs significantly from zero. If the frequency function of a correlation coefficient of this type could be found, it would be possible to test the hypothesis that the population correlation is zero and in this sense test the sequence for randomness."

This approach to the study of serial correlation, developed by Wald and Wolfowitz (10, pp. 378-388) and discussed by Hoel, will be followed here to test the thesis of the importance of serial correlation. Hoel (9, p. 343) shows that to approximate the serial correlation of lag one inherent in a sequence of numbers it is sufficient to consider the statistic

$$R = \sum_{i=1}^{n} (x_i)(x_{i+1}). \qquad (7)$$

Furthermore, as shown by Wald and Wolfowitz (10), the random variable R has an approximate normal distribution for large sample sizes "if it is assumed that the value of the sequence being tested constitute a random sample from a population that possesses low order moments" (9, p. 344). From basic statistics it can be shown that the mean, the variance, and other low order moments certainly exist for the theoretically uniform distribution from which the pseudo-random numbers are drawn.

Thus, we can establish zero serial correlation as the null hypothesis for the pseudo-random sequence. With the expected value and variance of R given as (9, p. 344):

$$E(R) = (S_1^2 - S_2)/(n - 1) \tag{8}$$

$$S_R^2 = \frac{S_2^2 - S_4}{n - 1} + S_1^4 - \frac{4S_1^2 S_2 + 4S_1 S_3}{(n - 1)(n - 2)}$$

$$\frac{S_2^2 - 2S_4}{(n - 1)(n - 2)} - E^2(R) \tag{9}$$

Where

$$S_1 = x_2 + x_2 + \ldots\ldots + x_n$$

$$S_2 = x_1^2 + x_2^2 + \ldots\ldots + x_n^2$$

$$S_3 = x_1^3 + x_2^3 + \ldots\ldots + x_n^3$$

$$S_4 = x_1^4 + x_2^4 + \ldots\ldots + x_n^4 \tag{10}$$

the statistical significance of the serial correlation indicated by the statistic R can easily be determined. Formulas (7) to (10) are applicable to serial correlation of any lag for which the corresponding value of R has been calculated.

The advantage of the above time series approach is that it is now possible to compute the statistical significance of positive or negative serial correlation of any lag in a partial sampling from a sequence of pseudo-random numbers.[2]

Although the main interest in the development of the thesis is serial correlation, a frequency test is also used. The objective of the frequency test is to test for goodness of fit with the uniform distribution by a two-tailed chi square analysis. The frequency test is standard in pseudo-random number generator analysis, and it is used here to obtain some additional information about the generators which are used in the sample simulation. Once the frequency information is obtained, the effects of this property upon the results of the simulation can be separated from the effects of serial correlation by additional statistical testing. The computer program for the frequency test was written by James W. Muir in conjunction with his Master's Thesis at Lehigh University in 1965 entitled "The Effect of Scheduling on a Multi-Variate Stochastic System Evaluated Using Simulation Techniques".

1.3 The Monte Carlo Problem

It remains now to choose a simulation experiment with which to test the effects of serial correlation in mixed generators. The technique to be employed is an extension of the Von Neumann and Ulam procedure of finding a probabilistic analogue to a mathematical problem and then obtaining approximate answers to the analogue by some experimental sampling procedure (12, pp. 2-3). The actual experiment used (13, pp. 1-8) is the evaluation of the normalized volume of an n-dimensional hypersphere, H_n, defined by

$$x_1^2 + x_2^2 + \ldots + x_n^2 \leq 1. \qquad (11)$$

The volume of the hypersphere, V_n, is given by

$$V_n = (2\pi^{n/2})/(n\Gamma(n/2)). \qquad (12)$$

If the hypersphere is surrounded by a cube of n dimensions, this hypercube, E_n, is defined by

$$-1 \leq X_i \leq 1, \quad i = 1,2,\ldots,n. \qquad (13)$$

Then the fraction of E_n contained in H_n is the normalized volume

$$v_n = (\pi^{n/2})/n2^{n-1}\Gamma(n/2)). \qquad (14)$$

This normalized volume may be computed by generating a pseudo-random sequence of points which lie in E_n and finding the fraction of points which lie in H_n called the success ratio.

The advantages inherent in the Monte Carlo integration for use as a sample simulation problem are the ease of computation and the ability to extend the problem to higher dimensions. The extension to higher dimensionality is desired because the number of dimensions, n, corresponds exactly to the number of pseudo-random numbers used each time through a simulation. It is not the purpose here to compare Monte Carlo with conventional methods of numerical mathematical analysis but simply to use the technique as a means of evaluating the effects of serial correlation in the pseudo-random sequences of mixed generators.

1.4 The Effects of Serial Correlation and Frequency Upon the Success Ratio

This thesis was started with several questions unanswered:

1. Can serial correlation be found and isolated in the sequences of mixed generators?

2. Once found, what will be the effects of serial correlation upon the sample simulations?

3. How will frequency properties affect the sample simulations?

The first question will be answered by the results of the experiment, and speculation need not be made beforehand. But it seems constructive to attempt premature answers to the second and third questions. In this way, the intuition gained from the literature can be compared with the actual results.

Let a highly significant positive serial correlation of lag one be the first assumed case. In this situation, each number would be very closely related to the next, and there would be very little variability within the generated sequence. If the sequence of numbers were all too high, the mean of the success ratios would be too low because too many points would fall outside of the hypersphere. Were the sequence of numbers all too low, the success ratio would appear too large because too few points would be found outside of the hypersphere. In both cases, the variance of the success ratios would be too small because of the decreased variability of the pseudo-random sequence.

The results of a highly significant negative serial correlation of lag one are more difficult to visualize, but it seems likely that the opposite of the above would occurr. If the first number in the sequence were a high one the success ratio would be too high and vice-versa. The variance of the success ratios should in this case be too large because of the increased variability within the pseudo-random sequence.

A significant serial correlation of a higher lag, say six, is not expected to have as strong an effect upon the success ratios, but when correlation of several lags occurrs at the same time the effect may be very pronounced.

In the development of this thesis it is also expected that significant frequency properties can bias the simulation results. The two cases of significant frequency are too much uniformity and non-uniformity. When a pseudo-random sequence is too uniform or fits too closely with the uniform distribution, it is expected that the mean of the normalized volumes should be correct but that the variance should be significantly small. The mean of the simulation results will be more correct than usual because a perfectly uniform distribution is the ideally random case in any success-failure situation. The variance is expected to be quite small because sequences which are too-uniform are all approaching the same perfectly uniform distribution.

A poor goodness of fit to the uniform distribution by a pseudo-random sequence, however, should yield a very high variance of the success ratios. This is expected because the non-uniformity has no set pattern within the sequences of numbers, and the results are bound to vary more than expected. It is quite difficult to predict the bias in the means of the success ratios due to non-uniformity except to say that the results should be either too large or too small depending upon the degree and the type of non-uniformity.

CHAPTER 2: DESIGN OF THE EXPERIMENT

The experiment designed to test the importance of serial correlation in the results of the Monte Carlo integration is divided into five parts. The purpose of the first part is to isolate mixed generators exhibiting statistically significant serial correlation characteristics. Part two is designed to test the frequency characteristics of the same generators. Part three uses generators with and without serial correlation in the Monte Carlo integration. The fourth part of the experiment tests the significance of the different generators. Finally, an assessment is made of the significance of the effects of serial correlation and frequency upon the results of the simulation.

The generators chosen for the experiment are limited to ones similar to those studied by Hull and Dobell (2) and others since some conclusions have already been drawn on the sequences obtained from these generators. Specifically, the parameters used are: $c=1$ and 337; $s=2,3,\ldots,13$; $m=2^{19}$. The values of c and s are in the range of generators previously cited, and m is determined by the word length of the GE 225 which is 19 bits plus a sign bit. The different generators are obtained by modifying a program written by S. Kranes of the GE Research Laboratory in Schenectady, N.Y..The starting value is chosen by the user, and zero is used exclusively in the experiment. In Appendix II is shown the original generator program, Program I, with the parameters $c=337$ and $s=8$.

All computer programs used in the development of this thesis can be found in the author's personal file.

2.1 Serial Correlation Analysis

For each of the generators:

$$x_{n+1} \equiv (2^s + 1)x_n + c \quad (\text{mod } 2^{19})$$

with $c = 1$, $s = 2,3,\ldots,13$

$c = 337$, $s = 2,3,\ldots,13$

the following calculations are performed as outlined on Flow Chart I in Appendix I.

1. Calculate S_1, S_2, S_3, S_4 by equations (10) for each of 32 groups of 512 numbers each.
2. Sum the values to obtain each of S_1, S_2, S_3, S_4 for three samples each of 512, 1024, 2048, 4096 numbers. i.e. Obtain the S values for each of the first three groups of 512 numbers, each of the first three groups of 1024 numbers, etc.. This is possible as values of S are additive for groups of numbers.
3. Calculate R as in equation (7) for each of 32 groups of 512 numbers for lags 1 to 15.[3]
4. Sum the values to obtain R for three samples each of 512, 1024, 2048, 4096 numbers. R is also additive for groups of numbers.
5. Calculate E(R) as in equation (8) for each sample of each size for each lag.
6. Calculate S_R as in equation (9) for each sample of each size for each lag.
7. Calculate T, the number of standard deviations of significance of R on the normal distribution by

$$T = (R - E(R))/S_R \tag{15}$$

for each sample of each size for each lag.

The group sizes of 512, 1024, 2048, 4096 were chosen because of their use by Davis and Rabinowitz (13) in their Monte Carlo experiment with n-dimensional spheres. Three samples of each group size are used in order that serial correlation might be verified over several samples instead of basing conclusions on a single result which could be purely local. A group size of 8192, next in the sequence, was attempted in the calculation of serial correlation, but it overextended the memory capabilities of the GE 225 for the computation techniques employed. Hence, with a sample size of three, the maximum total sample is three groups of 4096 numbers each.

The null hypothesis is that there is zero correlation in the sequence of numbers. Significant positive or negative correlation will suffice to reject the null hypothesis if it occurrs in each of the three samples tested for one group size. A two-tailed test on the standard normal distribution is employed with significance limits on T of $\pm$ 1.96 standard deviations for the .05 level of confidence and $\pm$ 2.58 standard deviations for the .01 level of confidence.

2.2 Frequency Test

The frequency test, as programed by Muir, uses group sizes of 500 numbers and establishes 100 frequency cells. This is done by examining the two leading digits of each number and placing it in the proper position in a ten by ten matrix. The expected value in each cell is five numbers for a uniform distribution in which each number has an equal probability of appearance. The observed frequency in the cells is tested against the expected frequency by the chi square distribution where for one group of 500 numbers

$$X^2 = \sum_{j=1}^{100} (\text{OBSERVED} - 5)^2/5 ,$$

and for y groups of 500 numbers

$$X^2 = \sum_{j=1}^{100} (\text{OBSERVED} - 5y)^2/5y .$$

The summation leaves the chi square statistic with 99, or approximately 100, degrees of freedom. It is desired to test for statistically significant uniformity as well as statistically significant non-uniformity so both tails of the chi square distribution are used in the test. The chi square values of significance at the .95 and .99 levels are 124.3 and 135.8. Above these limits a generator has non-uniform frequency characteristics. The values at .05 and .01 are 77.9 and 70.1. Below these limits a generator has frequency characteristics which are too uniform. Thus, we have upper and lower limits of significance at two levels of confidence for the chi square values obtained. The frequency test is performed for each generator of the first part of the experiment for 40 groups of 500 numbers each.

2.3 Monte Carlo Integration

After the separation of generators which exhibit serial correlation from those which do not, several of these generators can be used for the Monte Carlo integration. The selection of the evaluation of the normalized volumes of n-dimensional hyperspheres as the simulation problems for this thesis is based partially upon the ease of programming. Since the hypersphere is defined as in equation (11) by

$$x_1^2 + x_2^2 + \ldots + x_n^2 \leq 1$$

where n is the dimensionality, all that need be done is to generate n random numbers and inquire by computer whether the sum of the squares is greater than one. If so, the point generated lies outside the sphere but within the hypercube defined by $-1 \leq x_i \leq 1$ and is classified as a failure. If, on the other hand, the sum of the squares is greater than or equal to one the point must lie within or on the surface of the hypersphere and is classified as a success.

The above process is carried out for 32 groups of 512 numbers each with n varying from 2 to 15. The success ratios, the number of successes divided by the number of trials, are computed for each group. Once again, the groups are combined into three samples each of 512, 1024, 2048, 4096 numbers; success ratios are computed for each sample of each group size. These calculations are outlined on Flow Chart II in Appendix I.

Smaller samples are not tested here because rarely will less than 500 pseudo-random numbers be used in a computer simulation. A sample of 4096 numbers is sufficient for the purposes of this experiment as the success ratio begins to converge upon the correct normalized volume and the accuracy of the computations can thus be tested.

2.4 Significance of Integration Results

It may logically be asked at this point how it will be shown that one generator produces different results than another in the Monte Carlo integration, especially since the correct answers will not be realized in any case. The procedure to test for significance of results is based upon the fact that the success ratio is a binomially distributed random variable, and for the large sample sizes employed the binomial is a very close approximation to the normal distribution.

Let p_1, p_2, p_3 be the success ratios obtained from the three samples for a given generator, given dimensionality, and given group size. Let $\overline{p}$, then, be $(p_1, p_2, p_3)/3$. If u is the correct normalized volume for the given conditions, and if N is the group size, the theoretical variance of the average of the three samples is given by

$$s'_p{}^2 = (u(1 - u))/3N. \qquad (16)$$

The actual variance, calculated in the conventional manner with a correction for small samples, is

$$s_p{}^2 = \sum_{i=1}^{N} ((pi - \overline{p})^2)(N/(N - 1)). \qquad (17)$$

To test, then, for difference in variance between the estimate of the normalized volume and the correct result, the chi square statistic

$$X^2 = ((j - 1)s_p{}^2)/s'_p{}^2 \qquad (18)$$

where j is the number of samples and is equal to three, is calculated. This statistic has j - 1 or two degrees of freedom and for the two sided alternative the null hypothesis of no difference is rejected if

$$X^2_{.025,2} \geq 7.378 \ , \ X^2_{.975,2} \leq .0506 \ ,$$

for the .05 level or

$$X^2_{.005,2} \geq 10.597 \ , \ X^2_{.995,2} \leq .0010 \ ,$$

for the .01 level of confidence.

A significant difference between the mean of the estimates and the correct result is determined by calculating the standardized normal deviate

$$z = (\overline{p} - u)/s'_p . \tag{19}$$

The limits of significance are the same as for the normally distributed values of T which represent the significance of the serial correlation, i.e. for the .05 level $\pm$ 1.96 standard deviations and for the .01 level $\pm$ 2.58 standard deviations.

The significance of the results of the Monte Carlo integration is determined only for the largest sample size, 4096, for each dimensionality of each generator used. It is felt that the estimates by smaller sample sizes differ too greatly from the correct result to allow meaningful comparison between generators. The computations of this part of the experiment are routine and need not be described in the Appendix.

2.5 Significance of Serial Correlation, Frequency, and Dimensionality in Integration Results

The previous section, Significance of Integration Results, is designed to test the correctness of the estimated normalized volumes obtained from the generators used in the Monte Carlo integrations. This section, however, is set up to evaluate and separate the effects of the serial correlation and frequency properties of the different generators, as well as the effect of dimensionality, upon the results of the simulations.

Lehigh University's Analysis of Variance Library Program, D 3.005, for the GE 225 was used for the computations. Variations of the table below are used for the analyses of this section.

FACTOR A	FACTOR B	FACTOR C or REPLICATES						
Serial Correlation	Frequency	Dimensionality						
		2	3	4	5	6	7	8
Non-significant	Too-uniform							
	Non-significant							
	Non-uniform							
Significant	Too-uniform							
	Non-significant							
	Non-uniform							

The averages of the estimated normalized volumes, $\overline{p}$, are used as data in the columns under dimensionality. Classification of mixed generators by the factors serial correlation and frequency forms the rows of the table, and dimensionality forms the columns. Estimates of normalized volumes for dimensionalities of only two to eight are used because there exist too many missing values and because of the poor accurracy of the estimates at higher dimensionalities.

Three separate analyses are run from the above table as follows:

1. Serial correlation, frequency, and dimensionality are all considered input variables or factors. One replicate is obtained for each combination of variables. For example, non-significant serial correlation, too-uniform frequency, and dimensionality two constitute one experimental piece of data of which there are no further replications. Dimensionality eight is not used in the first or second analysis of variance because of the missing data point.

In this analysis the serial correlation mean square, A, has one degree of freedom; the frequency mean square, B, has two degrees of freedom; and the dimensionality mean square, C, has five degrees of freedom. The interactions AB, AC, and BC have two, five, and ten degrees of freedom respectively, and the residual has the remaining ten degrees of freedom. The F values of significance at the .05 and .01 levels of confidence from any table of the F distribution are as follows:

	F	
Variables	.05	.01
Serial Correlation (A)	4.96	10.00
Frequency (B)	4.10	7.56
Dimensionality (C)	3.33	5.64

	F	
Interactions	.05	.01
AB	4.10	7.56
AC	3.33	5.64
BC	2.98	4.85

The input variable or interaction will be considered significant if its resulting mean square divided by the residual mean square exceeds the F value for that variable or interaction.

2. In the second analysis, serial correlation and frequency are considered input variables, and dimensionalities two to seven are considered replications. In this way, the effects of serial correlation and frequency can better be isolated. For example, non-significant serial correlation, too-uniform frequency and dimensionalities two to seven constitute six replications of the same combination of input variables.

In this analysis serial correlation, frequency, and the residual mean square have one, two, and two degrees of freedom respectively. The F values of significance are:

	F	
Variables	.05	.01
Serial Correlation	18.50	98.50
Frequency	19.00	99.00

3. In case consideration of the entire set of data in a single analysis of variance would obscure the significant variables, a final analysis is performed with each dimensionality a separate case. For dimensionality two, for example, the six

values used in the analysis are the six combinations of serial correlation and frequency properties of the mixed generators tested. Dimensionality eight is included in this set of analyses of variance because of the ease of application of the missing value computations. The number of degrees of freedom and the F values are the same as in analysis number two.

CHAPTER 3: RESULTS

3.1 Serial Correlation and Frequency Results

Sample outputs for this portion of the experiment are shown in Appendix II for the generator c = 337 and s = 8. The calculations were spot checked by hand and by supplemental computer programs and all results, even those showing obvious repetition of values, appear to be accurate. Below is a description of the serial correlation and frequency characteristics of each generator tested. In this description, only significant positive or negative serial correlation results are reported; frequency results are reported whether or not they are significant. Note that lag refers to the lag of the serial correlation, sign refers to positive or negative correlation, level refers to the degree of confidence, group refers to the size of the sample (Group 1 = 512 numbers, Group 2 = 1024 numbers, Group 3 = 2048 numbers, Group 4 = 4096 numbers), and T refers to the number of standard deviations on the normal distribution. The serial correlation is considered present only if significant positive or negative correlation is apparent in each of the three samples of a group.

Table I: SERIAL CORRELATION AND FREQUENCY RESULTS

Generator	Lag	Sign	Level	Group	T		
c=337,s=2	1		05,01	1	4.605	4.768	5.732
				2	6.595	8.099	6.393
				3	10.500	9.194	9.924
				4	13.912	13.934	11.734
	2		05	4	2.253	2.556	2.276

Frequency test not significant: $X^2 =$ 94.88

Generator	Lag	Sign	Level	Group	T		
c=337,s=3	1		05,01	2	2.792	3.964	3.361
				3	4.785	4.113	5.652
				4	6.276	7.683	6.955

Frequency test not significant: $X^2 =$ 92.79

Generator	Lag	Sign	Level	Group	T		
c=337,s=4	1		05,01	4	3.273	3.485	3.296

Frequency test not significant: $X^2 =$ 105.06

c=337,s=5

Correlation not significant

Frequency test significant at .05, not significant at .01:

$X^2 =$ 71.87

c=337 s=6

Correlation not significant

Frequency test significant at .05, .01: $X^2 =$ 170.09

c=337,s=7

Correlation not significant

Frequency test significant at .05, not significant at .01:

$X^2 =$ 75.68

c=337,s=8

Correlation not significant

Frequency test significant at .05, .01: $X^2 =$ 65.10

Generator	Lag	Sign Level	Group	T		

c=337,s=9

Correlation not significant

Frequency test not significant: $X^2 = 79.24$

c=337,s=10

Correlation not significant

Frequency test significant at .05, .01: $X^2 = 47.74$

c=337,s=11

Correlation not significant

Frequency test not significant: $X^2 = 123.96$

c=337,s=12

Correlation not significant

Frequency test significant at .05, .01: $X^2 = 17.30$

c=337,s=13

Correlation not significant

Frequency test significant at .05, .01: $X^2 = 15.33$

Generator	Lag	Sign Level	Group	T		
c=1,s=2	1	05,01	1	3.275	5.138	4.111
			2	5.979	6.104	7.122
			3	8.607	10.594	9.990
			4	13.591	13.925	11.367
	2	05	4	3.671	1.759	3.291

Frequency test not significant: $X^2 = 107.95$

Generator	Lag	Sign Level	Group	T		
c=1,s=3	1	05	1	2.150	2.962	2.791
		05,01	3	5.937	2.861	5.187
			4	6.195	6.921	8.546

Frequency test not significant: $X^2 = 98.53$

Generator	Lag	Sign	Level	Group	T		
c=1,s=4	1		05	2	2.631	3.404	1.974
			05,01	3	4.230	3.274	3.876
				4	5.297	4.488	3.459

Frequency test not significant: $X^2 = 96.94$

c=1,s=5	1		05	4	2.206	2.919	1.986

Frequency test significant at .05, .01: $X^2 = 69.28$

c=1,s=6

Correlation not significant

Frequency test significant at .05, .01: $X^2 = 65.53$

c=1,s=7	4		05,01	1	3.377	2.725	3.642

Frequency test significant at .05, .01: $X^2 = 36.21$

c=1,s=8	2	-	05	4	1.967	1.967	1.967
	4	-	05,01	1	2.503	3.045	3.012
				2	3.952	3.920	3.952
				3	5.585	5.585	5.585
				4	7.914	7.914	7.914
	8	-	05,01	1	10.901	11.556	11.078
				2	15.903	15.765	15.903
				3	22.407	22.407	22.407
				4	31.702	31.702	31.702
	12	-	05	1	3.070	2.571	4.018
			05,01	2	4.018	3.282	4.018
				3	5.179	5.179	5.179
				4	7.340	7.340	7.340

Frequency test significant at .05, .01: $X^2 = 13.22$

Note: The same values of S were obtained for this generator for groups: 1, 5, 9, 13, 17, 21, 25, 29; 2, 6, 10, 14, 18, 22, 26, 30; etc. The same values of R were obtained for each lag for each of the groups above.

c=1,s=9

Correlation not significant

Frequency test significant at .05, .01: $X^2 = 13.48$

Generator	Lag	Sign	Level	Group	T		
c=1,s=10							

Correlation not significant

Frequency test significant at .05, .01: $X^2 = 2.68$

Note: The same values of S were obtained for this generator for groups: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31; 2, 4, 6, etc. Values of R do not show repetition.

Generator	Lag	Sign	Level	Group	T		
c=1,s=11	11		05	1	1.998	2.558	2.558
			05,01	2	3.189	3.584	3.230
				3	4.767	4.294	3.291
				4	6.391	5.055	5.219
	13		05	1	1.775	2.135	2.135
			05,01	2	2.773	2.988	3.243
				3	4.022	4.743	5.465
				4	6.182	8.222	7.636
	14	-	05	1	2.284	2.035	2.035
			05,01	2	3.083	2.908	2.731
				3	4.257	3.760	3.263
				4	5.684	4.278	2.871

Frequency test significant at .05, .01: $X^2 = 2.34$

Note: The same values of S were obtained for this generator for the same groups as with c = 1, s = 8. Values of R do not show repetition.

Generator	Lag	Sign	Level	Group	T		
c=1,s=12	5		05,01	1	3.454	3.454	3.454
				2	4.851	4.851	4.562
				3	6.838	6.219	6.009
				4	9.217	8.004	6.872
	7		05	2	1.968	1.968	2.898
			05,01	3	2.761	4.728	5.380
				4	5.286	6.733	6.687
	9		05	4	1.531	3.735	3.688
	10		05	3	2.075	2.744	3.003
			05,01	4	3.413	4.726	6.039
	11		05	3	2.206	2.377	2.432
			05,01	4	3.221	3.549	3.877
	14	-	05,01	1	2.710	2.710	2.710
				2	3.862	3.862	3.154
				3	5.482	3.987	3.493
	15		05	4	6.710	6.147	9.335
			05,01	3	2.037	2.810	3.065
				4	3.412	4.866	6.319

Frequency test significant at .05, .01: $X^2 =$ 2.16

Note: The same values of S were obtained for this generator for groups: 1, 9, 17, 25; 2, 10, 18, 26; etc. Values of R do not show repetition.

Generator	Lag	Sign	Level	Group	T		
c=1,s=13	3	-	05	1	1.978	1.978	1.978
			05,01	2	2.828	2.828	3.228
				3	4.020	4.895	5.214
				4	6.306	7.386	7.995
	4	-	05	1	2.128	2.128	2.128
			05,01	2	3.039	3.039	3.766
				3	4.318	5.889	6.441
				4	7.220	9.120	5.721
	5	-	05	1	2.111	2.111	2.111
			05,01	2	3.015	3.015	4.165
				3	4.285	6.750	7.601
				4	7.805	10.761	7.993
	6		05,01	1	4.144	4.144	4.144
				2	5.827	5.827	4.156
				3	8.217	4.652	3.442
				4	9.093	4.854	2.343
	7		05	3	2.746	2.109	3.754
	9		05,01	3	2.216	3.725	4.205
				4	4.196	5.934	4.759
	10	-	05,01	1	5,942	5.942	5.942
				2	8.431	8.431	7.093
				3	11.943	9.135	8.229
				4	14.903	11.649	9.655
	12	-	05	3	2.193	3.764	4.315
			05,01	4	4.215	6.115	7.216
	13		05,01	1	3.067	3.067	3.067
				2	4.304	4.304	5.406
				3	6.064	8.368	9.115
				4	10.198	12.876	14.512
	15		05,01	1	4.196	4.196	4.196
				2	5.900	5.900	4.370
				3	8.321	5.054	3.944
				4	9.451	5.564	7.764

Frequency test significant at .05, .01: $X^2 =$ 136.45

Note: The same values of S were obtained for this generator for groups: 1, 17; 2, 18; 3, 19; etc. Values of R do not show repetition.

With c=337, the three generators exhibiting significant serial correlation have frequency characteristics which are non-significant. However, the nine generators with c=1 which gave significant serial correlation show significantly too-uniform, significantly non-uniform, and non-significant frequency results. Although no definite conclusions can be drawn at this point in the experiment, it is revealing to construct a classification of the serial correlation and the frequency characteristics of the generators tested.

A. Serial Correlation Test not Significant;
Frequency Test not Significant:

c = 337; s = 9, 11

B. Serial Correlation Test Significant;
Frequency Test not Significant:

c = 337; s = 2, 3, 4
c = 1 ; s = 2, 3, 4

C. Serial Correlation Test not Significant;
Frequency Test Significantly Non-Uniform

c = 337; s = 6

D. Serial Correlation Test not Significant;
Frequency Test Significantly Too-Uniform

c = 337; s = 5, 7, 8, 10, 12, 13
c = 1 ; s = 6, 9, 10

E. Serial Correlation Test Significant;
Frequency Test Significantly Non-Uniform

c = 1; s = 13

F. Serial Correlation Test Significant;
Frequency Test Significantly Too-Uniform

c = 1; s = 5, 7, 8, 11, 12

Mention must be made here of the peculiar phenomenon of the repetition of the values of S for the generators with c = 1 and s = 8, 10, 11, 12, 13 and of the repetition of the values of R for the generator with c = 1 and s = 8. Repetition of the values of R and S appears to be related both to significant serial correlation and to significantly too-uniform results. This holds true except for the generator c = 1, s = 10 which does not exhibit significant serial correlation and the generator c = 1, s = 12 which shows a significantly high frequency of occurrence. Also worthy of mention is the repetition in many instances of the final three digits of the S values, even though the first three digits may show variation.

The repetition of the values of R and S may be equivalent to the sub periods or harmonics of length 2^i discussed by Rotenberg (4, p. 611). The repetition of digits may indicate too little change in the low order digits of the generated numbers. An over stability of this type would not be exposed by the frequency test since only the two leading digits are tested and tallied.

The serial and frequency results generally support the statements made by Hull and Dobell (2, p. 36, 39) that small values of a give unsatisfactory serial results and that more complicated values of c tend to improve the performance of a generator. Serial correlation appears only for s = 2, 3, 4 with c = 337 in accord with the above conclusion regarding small values of a. The correlation in these generators is only as high as lag two. For c = 1 correlation only as high as lag two is revealed for s = 2, 3, 4, 5; but correlation of both low and high order appears for c = 1 and s = 7, 8, 11, 12, 13. The statement by Hull and Dobell can be modified here to say that small values of a can result in serial correlation of a lag less than or equal to two, whereas higher values of a can result in serial correlation of order greater than two.

With c = 337, only three out of twelve generators tested show significant serial correlation as compared to nine out of twelve generators tested with c = 1; and high values of a do not result in serial correlation with c = 337 as they do with c = 1. Thus is shown better performance for more complicated values of c.

3.2 Monte Carlo Integration Results

The generators chosen for the Monte Carlo integration are representative of the serial correlation and frequency characteristics of the generators tested in the previous section. Extreme cases are used when possible to bring out most strongly the effects of the properties upon the simulation. The generators used in the simulation are listed below with a summary of their properties.

1. c = 337, s = 9: No significant correlation; Frequency test not significant, $X^2 = 79.24$
2. c = 337, s = 6: No significant correlation; Frequency test non-uniform, $X^2 = 170.09$
3. c = 337, s = 13: No significant correlation; Frequency test too-uniform, $X^2 = 17.30$
4. c = 337, s = 3: Correlation of lag 1; Frequency test not significant, $X^2 = 92.79$
5. c = 1, s = 13: Correlation of lags 3, 4 5, 6, 7, 9, 10, 12, 13, 15; Frequency test non-uniform, $X^2 = 136.45$
6. c = 1, s = 11: Correlation of lags 11, 13, 14; Frequency test too-uniform, $X^2 = 2.34$

After obtaining the success ratios for each generator an additional simulation was run with

7. c = 337, s = 11: No significant correlation; Frequency test not significant, $X^2 = 123.96$

in order to clarify the results obtained.

The Monte Carlo simulation was carried out by computer for each of the generators listed above as described in the experimental design. A sample output of success ratios is shown in Appendix II for the generator c = 337, s = 8.

As a summary of the integration results, the means and standard deviations of the three estimates of the normalized volumes for the final group size of each dimensionality and each generator are presented below in Table II. Listed above the estimates obtained at each dimensionality are the correct normalized volumes calculated from formula (14) and taken from (13, p. 3, 4) and the theoretical standard deviations calculated from formula (16). Results for dimensionalities above twelve have been disregarded because of their inaccuracy.

Table II: MONTE CARLO INTEGRATION RESULTS

	$\bar{p}$	s'	$\bar{p}$	s'	$\bar{p}$	s'
	n=2		n=3		n=4	
	.785398	.003704	.523599	.004506	.308425	.004166
c=337,s=3	.781901	.004641	.533382	.002567	.338542	.003430
c=337,s=6	.777669	.004377	.519006	.006973	.293945	.010742
c=337,s=9	.785807	.023235	.526316	.017406	.301107	.038352
c=337,s=11	.787923	.007248	.543373	.016187	.313477	.004475
c=337,s=13	.786459	.001974	.520468	.004477	.311198	.002458
c=1,s=11	.744792	.008028	.522417	.002350	.283203	.007627
c=1,s=13	.831380	.002861	.598197	.002111	.286458	.005967

	n=5		n=6		n=7	
	.164493	.003344	.080746	.002458	.036912	.001701
c=337,s=3	.180016	.005735	.095930	.009531	.046734	.005430
c=337,s=6	.157363	.016161	.080911	.007170	.036599	.005932
c=337,s=9	.159385	.021379	.067345	.018173	.039414	.014758
c=337,s=11	.161812	.023228	.069283	.015541	.016329	.005160
c=337,s=13	.182039	.009633	.077035	.002518	.037162	.002925
c=1,s=11	.122573	.--1213	.073643	.002220	.048986	.003378
c=1,s=13	.140372	.005052	.033430	.017136	.004505	.007802

	$\bar{p}$	s'	$\bar{p}$	s'	$\bar{p}$	s'
	n=8		n=9		n=10	
	.015584	.001127	.006442	.000722	.002490	.000450
c=337, s=3	.021484	.001953	.008041	.006331	.003205	.002776
c=337, s=6	.018229	.001128	.008772	.006579	.004808	.002404
c=337, s=9	.016927	.005967	.010965	.007907	.004007	.001388
c=337, s=11	.002604	.002255				
c=337, s=13	.009766	.001954	.008772	.002193	.001603	.001388
c=1, s=11	.031250	.000000	.033626	.001266	.026442	.002404
c=1, s=13						

	n=11		n=12	
	.000920	.000273	.000326	.000163
c=337, s=3	.001773	.001536		
c=337, s=6	.001773	.001536		
c=337, s=9				
c=337, s=11				
c=337, s=13				
c=1, s=11	.026596	.004606	.023256	.000000
c=1, s=13				

3.3 Significance of Integration Results

It remains now to assign some significance to the results of the Monte Carlo integrations presented in Table II. Recall that the statistical tests to be used next are designed to assess the significance of difference between the theoretical and actual variances and the correct and estimated normalized volumes with a null hypothesis of no difference. The results of the statistical tests for the generators used in the simulation are shown below in Table III. The word no in this table refers to no significant difference and the acceptance of the null hypothesis, whereas the word yes refers to a significant difference and the rejection of the null hypothesis.

Table III: SIGNIFICANCE OF INTEGRATION RESULTS

c=337 s=3	s^2		$\bar{p}$	
n	.05	.01	.05	.01
2	no	no	no	no
3	no	no	yes	no
4	no	no	yes	yes
5	no	no	yes	yes
6	yes	yes	yes	yes
7	yes	yes	yes	yes
8	no	no	yes	yes
9	yes	yes	yes	no
10	yes	yes	no	no
11	yes	yes	yes	yes

c=337 s=6	s^2		$\bar{p}$	
n	.05	.01	.05	.01
2	no	no	yes	no
3	no	no	no	no
4	yes	yes	yes	yes
5	yes	yes	yes	no
6	yes	yes	no	no
7	yes	yes	no	no
8	no	no	yes	no
9	yes	yes	yes	yes
10	yes	yes	yes	yes
11	yes	yes	yes	yes

c=337 s=9	s^2		$\bar{p}$	
n	.05	.01	.05	.01
2	yes	yes	no	no
3	yes	yes	no	no
4	yes	yes	no	no
5	yes	yes	no	no
6	yes	yes	yes	yes
7	yes	yes	no	no
8	yes	yes	no	no
9	yes	yes	yes	yes
10	yes	yes	yes	yes

c=337 s=11	s^2		$\bar{p}$	
n	.05	.01	.05	.01
2	yes	no	no	no
3	yes	yes	yes	yes
4	no	no	no	no
5	yes	yes	no	no
6	yes	yes	yes	yes
7	yes	yes	yes	yes
8	yes	no	yes	yes

c=337 s=13	s^2		$\bar{p}$	
n	.05	.01	.05	.01
2	no	no	no	no
3	no	no	no	no
4	no	no	no	no
5	yes	yes	yes	yes
6	no	no	no	no
7	no	no	no	no
8	no	no	yes	yes
9	yes	yes	yes	yes
10	yes	yes	yes	no

c=1 s=11	s^2		$\bar{p}$	
n	.05	.01	.05	.01
2	yes	no	yes	yes
3	no	no	no	no
4	no	no	yes	yes
5	no	no	yes	yes
6	no	no	yes	yes
7	yes	no	yes	yes
8	-	-	yes	yes
9	no	no	yes	yes
10	yes	yes	yes	yes
11	yes	yes	yes	yes
12	-	-	yes	yes

c=1 s=13	s^2		$\bar{p}$	
n	.05	.01	.05	.01
2	no	no	yes	yes
3	no	no	yes	yes
4	no	no	yes	yes
5	no	no	yes	yes
6	yes	yes	yes	yes
7	yes	yes	yes	yes

An examination of the above results adds no small amount of confusion to the question of the existence and the importance of serial correlation. One generator which exhibits neither serial correlation or frequency of statistical significance $c = 337$, $s = 9$, rejects the null hypothesis of no difference in variance at both levels for each dimensionality. The null hypothesis of no difference between means is accepted, on the other hand, at both levels for all dimensionalities except six, nine, and ten.

Because the above was expected to be a relatively good generator but yielded such poor results the generator $c = 337$, $s=11$ was also tested. This generator which also shows neither serial correlation or frequency properties, yielded estimates of the normalized volumes for only eight dimensions. The null hypothesis is accepted for both variance and means at both levels for dimensionality four and at .01 for dimensionality two. Rejection of the null hypothesis occurrs for either means or variances at all other dimensions. It appears therefore that a generator which exhibits neither significant serial correlation or significant frequency is not necessarily an acceptable generator for simulation.

Holding with non-significant frequency results but adding a significant serial correlation of lag one, as with the generator $c = 337$, $s=3$, gives simulation results similar to the generator $c = 337$, $s=11$ above. Estimates of the normalized volumes are obtained for dimensions up to eleven and the null hypothesis is accepted at both levels for dimensionality two and for the .01 level at dimensionality three. The null hypothesis is rejected for means or variances at all other dimensions.

When a generator exhibiting correlation of three lags as well as too-uniform frequency results, c = 1, s = 11, is used for the simulation, the null hypothesis is accepted only for dimensionality two. It is rejected at all other dimensions at both levels by the variance test. Removing the serial correlation but retaining too-uniform frequency as with the generator c = 337, s = 13 yields some gratifying results. The hypothesis of no difference are accepted for five of the nine dimensions which is the best result obtained thus far. The conclusion drawn is that too-uniform frequency characteristics tend to give acceptable results for the Monte Carlo integration but only when serial correlation (of several lags) is absent.

Looking now at a generator with significant correlation of seven lags and a high frequency result, c = 1, s = 13, it can be seen that the test for means rejects the null hypothesis at each of six dimensionalities. (This is contrasted with the generator c = 337, s = 9 which rejected the null hypothesis at each dimensionality on the basis of the variance test.) The variance test for c = 1 s = 13 accepts the null hypothesis at both levels for all but the last two dimensionalities. Retaining non-uniform frequency but removing serial correlation as with c = 337, s = 6, yields a slightly better generator with acceptance of the null hypothesis at both levels for dimensionality three and at the .01 level for dimensionalities two and eight. The null hypothesis is rejected for all other dimensions. These results are by no means correct as those of c = 337, s = 13 which has no serial correlation but too-uniform instead of non-uniform frequency. Again, the conclusion that an acceptable generator contains both too-uniform frequency and non-significant serial correlation is supported.

3.4 Results of Significance Tests on Serial Correlation, Frequency, and Dimensionality

The analyses of variance discussed in this section are designed to assess the significance of serial correlation, frequency, and dimensionality upon the integration results. The data for this section are taken from Section 3.2, Table II, Monte Carlo Integration Results. The only generator which is used for the simulation and which is not included in the analyses of variance is $c = 337$, $s = 9$. The six other generators are included in this section and their properties are outlined in Section 3.2.

The results of the analyses of variance are reported below in the same order as Section 2.5 of the Design of the Experiment:

1. Serial correlation, frequency, and dimensionality are the input variables for this analysis with one replication of the experiment, (Table IV). The results reveal that only dimensionality is significant with an F value of 544. Serial correlation, frequency, and the first order interactions of the three input variables are insignificant in affecting the simulation results.

2. For this analysis of variance, serial correlation and frequency are the input variables and the dimensionalities are the replications (Table V). In this case again, neither serial correlation or frequency is significant. The results of the first analysis are thus supported.

3. Table VI shows the analyses of this section in which each dimensionality is considered to be a separate case. For dimensionalities two to seven, both serial correlation and frequency are again non-significant while dimensionality eight shows both variables to be significant. Since for this dimensionality a missing value calculation is performed (14, p. 31), and since the variance of the normalized volumes becomes quite large at higher dimensionalities, this case will not be considered important in the conclusions drawn from the analyses of variance.

The analyses of this section are not to be confused with the significance tests of previous sections of this experiment. It is quite possible that a generator exhibiting significant serial correlation and frequency properties and yielding simulation results significantly different from the correct results would in this section show non-significant serial correlation and frequency effects upon the results of the simulations. These results must therefore be considered important in their own right in drawing conclusions for this thesis.

It can be concluded from the results of the analyses of variance of this section that significantly different serial correlation and frequency properties and their interactions are not statistically significant at the .05 and .01 levels of confidence in affecting the results of the Monte Carlo simulations.

Table IV: ANALYSIS OF VARIANCE WITH SERIAL CORRELATION FREQUENCY, AND DIMENSIONALITY AS VARIABLES

Serial Correlation (A)	Frequency (B)	Dimensionality (C)	Estimates of Normalized Volume
Non-significant	Too-uniform	2	.786459
		3	.520468
		4	.311198
		5	.182039
		6	.077035
		7	.037162
	Non-significant	2	.787923
		3	.543373
		4	.313477
		5	.161812
		6	.069283
		7	.016329
	Non-uniform	2	.777669
		3	.519006
		4	.293945
		5	.157363
		6	.080911
		7	.036599
Significant	Too-uniform	2	.744792
		3	.522417
		4	.283203
		5	.122573
		6	.073643
		7	.048986
	Non-significant	2	.781901
		3	.533382
		4	.338542
		5	.180016
		6	.095930
		7	.046734
	Non-uniform	2	.831380
		3	.598197
		4	.286458
		5	.140372
		6	.033430
		7	.004505

RESULTS: $F_A = 1.19 \times 10^{-3}$ $\quad F_{AB} = 1.23$

$F_B = 7.34 \times 10^{-1}$ $\quad F_{AC} = 4.14 \times 10^{-1}$

$F_C = 5.44 \times 10^{-2}$ $\quad F_{BC} = 6.67 \times 10^{-1}$

Table V: ANALYSIS OF VARIANCE WITH SERIAL CORRELATION AND FREQUENCY AS VARIABLES AND DIMENSIONALITY AS REPLICATES

Serial Correlation	Frequency	Dimensionality 2	3	4
Non-significant	Too-uniform	.786459	.520468	.311198
	Non-significant	.787923	.543373	.313477
	Non-uniform	.777669	.519006	.293945
Significant	Too-uniform	.744792	.522417	.283203
	Non-significant	.781901	.533382	.338542
	Non-uniform	.831380	.598197	.286458

Serial Correlation	Frequency	5	6	7
Non-significant	Too-uniform	.182039	.077035	.037162
	Non-significant	.161812	.069283	.016329
	Non-uniform	.157363	.080911	.036599
Significant	Too-uniform	.122573	.073643	.048986
	Non-significant	.180016	.095930	.046734
	Non-uniform	.140372	.033430	.004505

RESULTS: $F_A = 9.47 \times 10^{-2}$

$F_B = 5.98 \times 10^{-1}$

Table VI: ANALYSIS OF VARIANCE WITH SERIAL CORRELATION AND FREQUENCY AS VARIABLES AND EACH DIMENSIONALITY AS A SEPARATE CASE

Serial Correlation	Frequency	Dimensionality 8
Non-significant	Too-uniform	.009766
	Non-significant	.002604
	Non-uniform	.018229
Significant	Too-uniform	.031250
	Non-significant	.021484
	Non-uniform	.038411

The analysis of variance tables for dimensionalities two to seven are constructed from the columns of Table V.

RESULTS:

Dimensionality	F_A	F_B
2	3.77×10^{-3}	6.6×10^{-2}
3	7.19×10^{-1}	5.9×10^{-1}
4	5.06×10^{-2}	2.02
5	7.47×10^{-1}	3.72×10^{-1}
6	1.40×10^{-1}	4.93×10^{-1}
7	3.33×10^{-1}	5.00×10^{-1}
8	7.21×10^{2}	1.56×10^{2}

CHAPTER 4: CONCLUSIONS

Several statements can now be made concerning mixed congruential pseudo-random number generators in the light of the isolation and the quantification of the serial correlation and frequency properties and the results of the sample simulations. Two criteria of judgement have been established in the presentation of this thesis, and each must be considered separately in the conclusions which are drawn.

A. The first criterion is that of the agreement of the estimates of the normalized volumes of the hyperspheres with the correct normalized volumes. The conclusions which can be drawn on this basis are as follows:

1. With serial correlation absent in a mixed congruential pseudo-random number generator, frequency characteristics which are too-uniform produce more accurate results than frequency characteristics which are non-significant or non-uniform.

2. Next, it can be seen from the results of the Monte Carlo integrations that serial correlation can cause bias in a simulation. The evidence of this is the decrease in performance when serial correlation is added to generators of varying frequency properties.

3. A final conclusion to be drawn from the first criterion is that serial correlation of several lags of both low and high order (as in $c=1$, $s=13$) is necessary to add a great deal of bias to simulation results. And it must

be noted that there seems to be no relation between the lag of the correlation and the dimensionality of the biased result. Therefore, the effect of serial correlation in this experiment does not depend upon the number of numbers used each time through a simulation.

B. Second is the criterion of the significance of the effects of serial correlation and frequency upon the simulation results as determined by the analyses of variance. The conclusions are as follows:

1. For the generators tested under the first criterion, fewer significant differences between the estimated and the true means and variances are obtained when a generator does not possess serial correlation. However, the word "fewer" in the preceeding sentence does not mean significantly fewer. Under the second criterion, therefore, the statement can be made that from the results of the experiment the presence of significant serial correlation in a mixed generator does not have a significant effect upon the results of the sample simulations.

2. A similar conclusion can be drawn for the frequency properties of mixed generators. Although significantly too-uniform frequency properties yield more accurate simulation results, frequency is not statistically significant in affecting these results.

Care must be used in any generalization of the above conclusions since they depend very much upon the limited number of generators tested, upon the design of the statistical tests, and upon the Monte Carlo integration used as the sample simulation.

SUGGESTIONS FOR FURTHER STUDY

My suggestions for further study in the field of pseudo-random number generation and testing are as listed below:

1. To isolate and test serial correlation in different mixed generators and in generators of the multiplicative type.

2. To apply the analysis of this thesis to different simulations, possibly industrial in nature.

3. To widen the scope of this analysis by including additional statistical tests, especially for runs and for short periodic properties.

4. To study the effect of the design of the computer program upon the properties of the pseudo-random number generator.

5. To study in depth the effects upon a simulation of a pseudo-random sequence which exhibits frequency properties which are too uniform.

6. To carry out the computations of the theoretical serial correlation formulas developed in the literature and compare the results with those of the time series approach of this thesis.

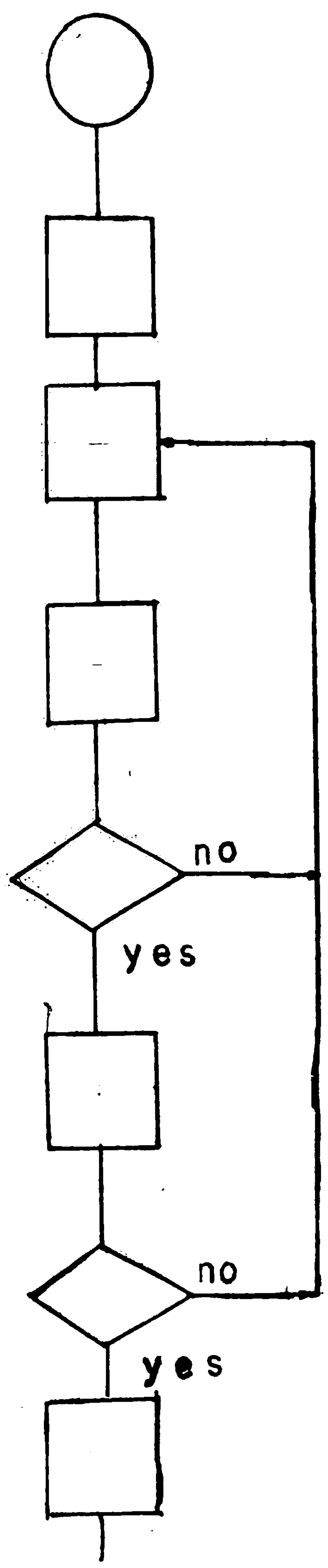
7. To study the relationship between serial correlation and runs in a pseudo-random sequence.

APPENDIX I

Flow Charts

FLOW CHART I: Calculation of $\bar{T}$

Symbol / Description



Start

Dimension and initialize

Generate a random number

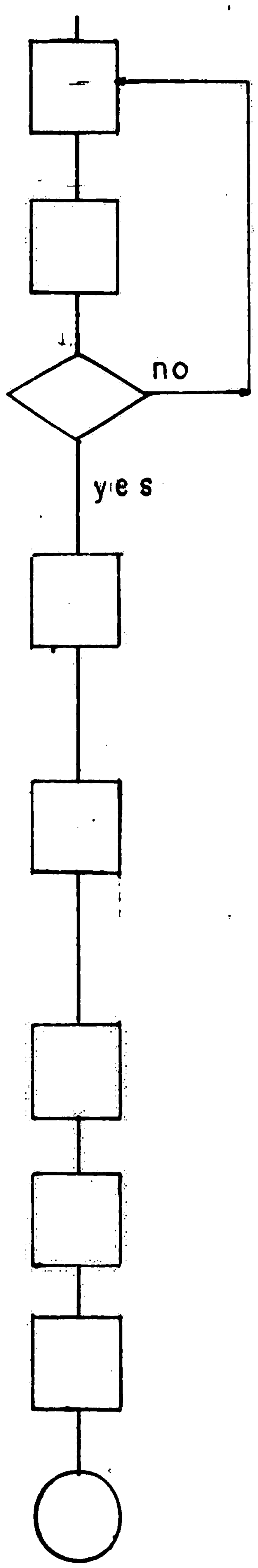
Calculate and accumulate S_1, S_2, S_3, S_4

Have calculations been performed for 512?

Store results of S_1, S_2, S_3, S_4 for group of 512 numbers

Have calculations been performed for 32 groups of 512 numbers?

Restart generator



Generate 527 random numbers and store in an array

Calculate R for lags one to fifteen for group of 512 numbers and store

Have R's been calculated for 32 groups of 512 numbers each?

Form S_1, S_2, S_3, S_4 into three samples each of 512, 1024 2048, 4096 numbers

Form R's into three samples each of 512, 1024, 2048, 4096 numbers

Calculate E(R), S(R) for each sample

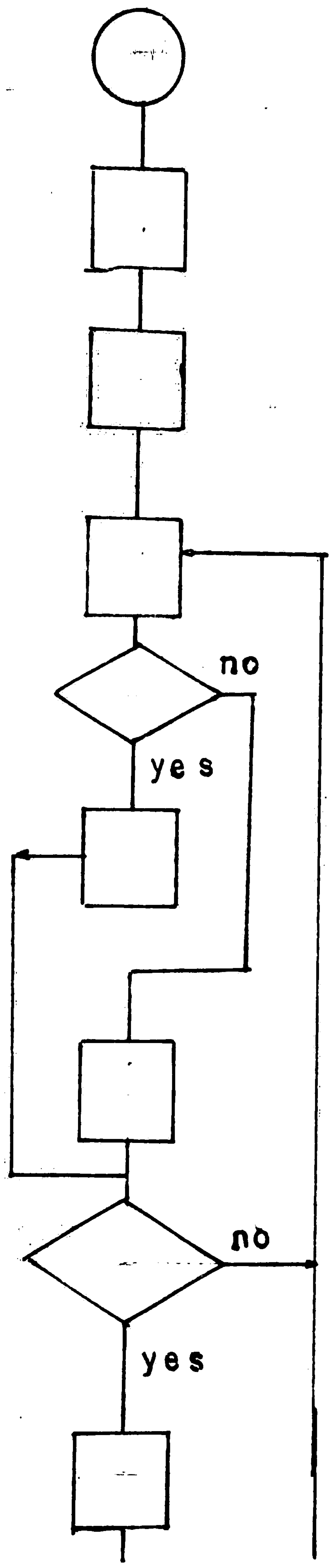
Calculate T for each sample

Print results

End

FLOW CHART II: Monte Carlo Simulation

Symbol



Description

Start

Dimension and initialize

Set dimensionality, N, equal to two

Generate N random numbers and calculate sum of squares

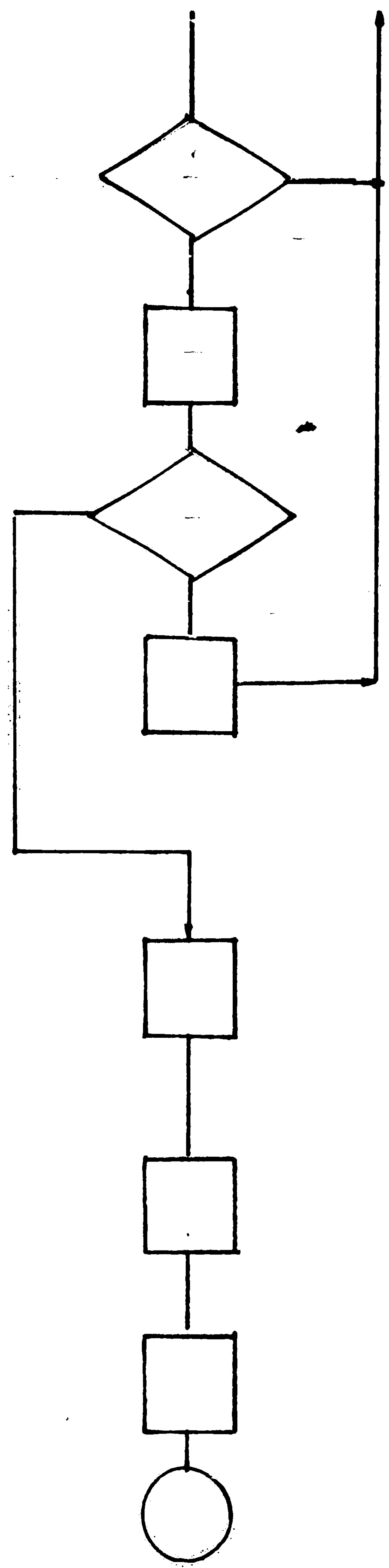
Is sum of squares 1?

Tally one success

Tally one failure

Have 512 numbers been used?

Store successes and failures for group of 512 numbers



Does number of groups of 512 numbers exceed 32?

Increment N by one

Does N exceed 15?

Restart generator

Form successes and failures into three samples each of 512, 1024, 2048, 4096 numbers

Calculate success ratio for each sample

Print results

End

PROGRAM I

Mixed Congruential Pseudo-Random Number Generator

with $c = 337$ and $s = 8$

```
                                                          GET D2.513  RANDOM NUMBER GENERATOR FLAT
                                                          00188 GRL 03FEB64N 225  VFAP  42
                                                          RANDOM NUMBER GENERATOR FLAT DISTRIBUTION
                                ETY      FLAT      RANDOM NUMBER GENERATOR - FLAT DISTRIBUTION
00000   0020001 0        FLAT   LDA      1,1       S. KRANES - G.E. RESEARCH LAB. SCHENECTADY
00001   2700022 2               STO      SENDU     METHOD USED - MIXED CONGRUENTIAL
00002   2700003 2               STO      GPL4      DATE    1/24/64
00003   0000000 0        GPL4   LDA      0
00004   2100041 2               CAB      ZERO
00005   2600036 2               BRU      GPL7
00006   2600010 2               BRU      RAN
00007   0300042 2               STA      U1
00010   0000042 2        RAN    LDA      U1
00011   2512010 0               SLA      8
00012   0100042 2               ADD      U1
00013   2514001 0               BMI
00014   2504040 0               CHS
00015   0100043 2               ADD      C
00016   2514001 0               BMI
00017   2504040 0               CHS
00020   2514002 0               BZE      RAN+1
00021   2600011 2
00022   0300000 0        SENDU  STA      0
00023   0300044 2               STA      II
00024   1000044 2               DLD      II
00025   2513023 0               NOR      19
00026   2511010 0               SRD      8
00027   1300046 2               DST      FLWD
00030   0000000 0               LDA      0
00031   0100050 2               ADD      CON
00032   2512013 0               SLA      11
00033   2300046 2               ORY      FLWD
00034   3000046 2               FLD      FLWD
00035   2620001 0               BRU      1,1
00036   0000051 2        GPL7   LDA      U
00037   0300042 2               STA      U1
00040   2600010 2               BRU      RAN
00041   0000000 0        ZERO   OCT      0
00042   1432771 0          U1   OCT      1432771
00043   0000337 0           C   OCT      337
          00044            II   BSS*     2
          00046          FLWD   BSS*     2
00050   3777755 0         CON   DEC      -19
00051   1432771 0           U   OCT      1432771
          00000                 END                LAST CARD  FLAT RANDOM NUMBER GENERATOR
```

SAMPLE OUTPUT I

Calculation of Significance of Serial Correlation

with c = 337 and s = 8

VALUES OF S

1	2	3	4	5	6	7	8
259.852	253.352	265.852	254.352	253.852	240.352	256.852	250.352
172.596	171.088	180.322	170.831	169.382	152.920	173.339	164.318
128.551	130.124	136.559	128.518	127.079	110.294	131.762	122.224
102.010	105.306	110.083	102.716	101.451	85.473	106.633	97.323
9	10	11	12	13	14	15	16
247.852	263.352	250.852	260.352	266.852	246.352	256.852	258.352
162.496	178.850	163.003	173.032	180.995	163.187	172.222	175.398
120.557	135.210	119.593	128.620	137.232	122.681	129.057	134.067
95.834	108.519	93.906	101.699	110.673	98.612	102.749	109.147
17	18	19	20	21	22	23	24
251.852	257.352	252.852	256.352	251.852	263.352	266.852	269.352
167.076	167.998	169.466	169.871	164.738	179.870	179.095	184.072
124.888	123.249	128.605	127.177	121.589	137.166	134.787	140.295
99.410	96.560	104.419	102.015	95.942	110.863	108.221	113.554
25	26	27	28	29	30	31	32
260.852	248.352	258.852	250.352	255.852	266.352	240.852	249.352
177.067	163.014	174.998	166.176	171.687	178.799	155.353	162.196
134.579	121.788	132.580	124.913	130.061	133.849	113.836	118.696
108.944	97.599	106.706	100.300	105.282	106.654	89.489	92.879

VALUES OF R

1	2	3	4	5	6	7	8	9	10
130.030	132.286	133.454	133.954	131.871	132.680	133.675	131.504	129.827	133.814
122.235	124.936	121.883	126.272	121.411	125.050	126.084	126.197	123.029	126.037
137.208	137.408	137.546	135.509	138.100	136.576	135.764	138.409	139.768	139.244
128.347	125.971	126.409	129.804	127.102	123.937	126.344	129.597	130.536	127.049
125.362	126.264	126.819	128.468	122.119	129.390	125.211	124.736	125.009	121.737
113.960	112.412	111.515	108.921	109.493	113.397	112.395	109.913	112.709	112.766
126.683	129.320	130.807	128.022	128.586	130.090	130.484	129.640	130.309	131.612
123.194	122.900	124.542	122.477	123.358	120.105	121.392	120.261	119.569	123.613
121.409	120.340	120.380	120.950	120.876	116.892	120.291	119.463	120.327	119.052
136.958	136.007	135.025	138.074	133.104	132.472	136.162	136.686	135.589	137.024
121.211	122.619	124.072	124.101	123.543	123.915	122.635	120.811	126.846	121.675
132.361	132.665	132.096	132.433	130.521	133.888	134.015	132.369	132.601	137.649
140.040	140.022	140.031	139.487	139.106	141.354	139.418	139.454	138.986	138.954
120.183	117.661	117.202	116.114	119.963	119.620	119.972	120.312	120.300	119.662
129.602	128.176	129.569	129.668	129.389	128.876	130.956	129.374	125.499	128.585
130.188	128.695	127.348	128.847	128.844	124.522	132.385	132.128	128.083	128.638
125.208	120.780	122.646	124.100	127.219	127.060	124.960	125.414	122.848	125.456
128.612	130.422	130.047	126.503	129.889	129.384	128.875	127.026	129.608	129.726
126.684	122.591	122.585	120.531	122.772	122.453	125.952	125.731	123.281	125.232
130.300	128.142	128.551	129.009	128.334	129.894	131.044	127.126	127.371	128.565
124.059	127.111	123.984	124.099	122.395	123.493	124.685	122.476	124.671	120.438
134.510	135.386	135.339	132.417	133.220	138.206	133.417	136.537	136.021	131.661
139.097	139.784	137.187	140.736	137.622	142.371	138.344	139.365	140.926	142.436
142.604	141.412	141.648	143.205	143.812	144.000	140.248	143.024	145.092	141.277
131.961	129.310	132.987	129.286	134.798	129.978	127.775	132.217	133.821	132.689
119.183	116.490	121.621	120.933	121.269	123.063	121.205	120.099	119.179	121.880
131.943	131.329	130.979	132.065	131.229	132.935	130.009	131.198	129.984	130.507
120.330	125.570	120.148	120.419	125.039	122.300	121.596	120.733	120.557	119.592
126.979	127.183	126.389	127.986	128.583	130.240	129.406	129.349	125.983	132.480
138.492	141.430	138.202	139.750	139.877	136.212	137.777	137.980	137.488	136.474
113.375	114.926	112.292	112.298	112.005	108.953	112.505	110.374	110.003	108.787
120.711	118.706	123.334	124.230	122.382	123.319	120.522	120.457	122.632	121.129

11	12	13	14	15
131.649	131.031	131.863	128.339	132.792
122.723	125.497	122.309	122.749	127.609
136.770	137.776	135.789	137.408	136.562
128.569	128.171	125.968	129.500	127.464
121.427	126.209	125.074	122.551	126.516
113.673	114.598	115.257	115.215	113.558
127.202	132.235	129.925	128.644	129.027
121.612	119.277	120.263	117.990	119.406
118.415	122.179	119.590	122.178	119.598
136.707	137.044	137.624	137.880	132.921
123.529	121.516	123.585	122.266	121.881
134.761	134.994	132.490	132.788	130.636
140.292	139.450	136.970	138.743	138.498
119.112	117.788	116.319	117.294	116.506
128.827	128.950	128.054	126.100	129.155
129.726	128.924	128.956	129.749	131.501
126.009	121.124	125.054	125.703	123.436
127.096	134.012	126.314	129.804	132.582
121.862	120.068	123.574	128.376	125.736
128.669	124.194	126.895	125.326	129.687
123.188	122.030	123.716	126.012	124.586
133.922	137.072	135.340	136.069	133.701
140.265	139.770	139.262	134.114	139.304
141.827	141.912	145.447	141.198	140.211
130.962	132.746	130.447	131.592	128.661
120.699	112.806	121.104	120.747	118.529
134.076	132.378	129.360	130.658	128.350
119.524	121.806	121.151	125.245	120.164
129.743	132.626	131.395	127.079	128.314
138.529	132.822	136.067	137.056	139.616
115.852	115.056	110.954	113.265	115.129
122.196	123.746	119.953	121.783	123.811

GROUPED VALUES OF S

1	2	3
259.852	253.352	265.852
172.596	171.088	180.322
128.551	130.124	136.559
102.010	105.306	110.083
513.204	520.204	494.204
343.683	351.153	322.302
258.675	265.077	237.373
207.316	212.799	186.924
1033.408	1001.408	1022.408
694.837	659.959	677.381
523.752	491.358	503.981
420.115	390.880	399.958
2034.816	2050.816	2069.816
1354.796	1369.184	1382.185
1015.110	1027.018	1037.756
810.995	821.138	830.985

GROUPED VALUES OF R

1	2	3
130.030	122.235	137.208
252.266	265.556	239.322
517.821	489.199	511.939
1007.021	1031.951	1051.075
132.286	124.936	137.408
257.221	263.379	238.676
520.600	490.896	511.631
1011.496	1026.184	1045.627
133.454	121.883	137.546
255.337	263.955	238.334
519.292	493.683	511.574
1012.975	1025.724	1041.986
133.954	126.272	135.509
260.225	265.313	237.389
525.539	487.888	515.558
1013.426	1029.673	1040.599
131.871	121.411	138.100
253.282	265.202	231.612
518.484	483.556	508.045
1002.040	1025.348	1045.263
132.680	125.050	136.576
257.730	260.514	242.786
518.243	492.982	507.168
1011.225	1021.540	1056.860
133.675	126.084	135.764
259.759	262.108	237.606
521.867	489.482	513.103
1011.348	1035.835	1047.527
131.504	126.197	138.409
257.701	268.007	234.649
525.707	484.549	509.328
1010.257	1030.596	1046.700
129.827	123.029	139.768
252.856	270.304	237.718
523.160	487.596	515.363
1010.756	1028.231	1049.818
133.814	126.037	139.244
259.851	266.293	234.503
526.144	489.728	515.399
1015.872	1031.238	1044.792
131.649	122.723	136.770
254.372	265.339	235.100
519.711	483.914	513.412
1003.624	1031.369	1042.839
131.031	125.497	137.776
256.528	265.947	240.807
522.474	492.318	515.732
1014.793	1030.844	1040.182
131.863	122.309	135.789
254.171	261.757	240.332
515.929	490.520	513.288
1006.449	1023.587	1045.602
128.339	122.749	137.408
251.088	266.908	237.766
517.996	484.399	515.112
1002.395	1026.998	1046.623
132.792	127.609	136.562
260.401	264.026	240.074
524.427	488.507	505.036

1012.934	1020.697	1049.242

EXPECTED VALUES OF R AND S

1	2	3
131.801	125.276	137.959
1.798	2.019	1.867
257.121	264.185	238.431
2.701	2.714	2.617
521.367	489.574	510.328
3.830	3.762	3.689
1010.775	1026.735	1045.850
5.373	5.349	5.253

THREE VALUES OF T FOR EACH GROUP SIZE FOR EACH LAG

1	2	3
-0.985	-1.506	-0.402
-1.797	0.505	0.340
-0.926	-0.100	0.437
-0.699	0.975	0.995
0.269	-0.169	-0.295
0.037	-0.297	0.093
-0.200	0.351	0.353
0.134	-0.103	-0.043
0.919	-1.680	-0.221
-0.660	-0.085	-0.037
-0.542	1.092	0.338
0.409	-0.189	-0.736
1.197	0.493	-1.312
1.149	0.416	-0.398
1.089	-0.448	1.418
0.493	0.549	-1.000
0.039	-1.914	0.076
-1.421	0.375	-2.605
-0.753	-1.599	-0.619
-1.626	-0.259	-0.112
0.489	-0.112	-0.740
0.225	-1.353	1.664
-0.815	0.906	-0.857
0.084	-0.971	2.096
1.042	0.400	-1.176
0.976	-0.765	-0.315
0.131	-0.025	0.752
0.107	1.701	0.319
-0.166	0.456	0.241
0.215	1.408	-1.445
1.133	-1.336	-0.271
-0.096	0.722	0.162
-1.098	-1.113	0.969
-1.579	2.255	-0.272
0.468	-0.526	1.365
-0.003	0.280	0.755
1.120	0.377	0.688
1.011	0.777	-1.501
1.247	0.041	1.375
0.949	0.842	-0.202
-0.085	-1.265	-0.637
-1.018	0.425	-1.273
-0.432	-1.504	0.836
-1.331	0.866	-0.573
-0.428	0.109	-0.098
-0.220	0.649	0.908
0.289	0.729	1.465
0.748	0.768	-1.079

0.034	-1.470	-1.162
-1.092	-0.894	0.726
-1.420	0.251	0.803
-0.805	-0.588	-0.047

-1.926	-1.252	-0.295
-2.233	1.003	-0.254
-0.880	-1.375	1.297
-1.560	0.049	0.147

0.551	1.155	-0.748
1.214	-0.059	0.628
0.799	-0.284	-1.435
0.402	-1.129	0.646

SAMPLE OUTPUT II

Calculation of Significance of Frequency

with $c = 337$ and $s = 8$

*****TEST OF RANDOM NUMBER GENERATOR*****

CONSTANTS* GROUP IS A BLOCK OF 500 CONCECUTIVE NUMBERS
NUMBER OF CLASS INTERVALS FOR GRAND FREQUENCY TEST IS 100

VARIABLES* NUMBER OF GROUPS TESTED = 40
SIZE OF SET TESTED FOR PERIODICITY = 10
NUMBER GENERATED PRIOR TO START OF TESTS = 0
STARTER NUMBER FOR GENERATOR = 0

GRAND FREQUENCY TABLE*   CELL NUMBER = 10*I + J   GRAND CHI SQUARE = 65.100000

I \ J	1	2	3	4	5	6	7	8	9	10
0	200	212	205	194	196	211	196	200	214	187
1	196	216	192	212	192	187	207	185	206	187
2	213	172	213	205	214	200	190	212	205	185
3	205	211	181	211	211	187	184	204	199	194
4	203	206	207	201	200	211	200	173	216	195
5	208	195	179	211	195	194	183	211	180	205
6	223	200	191	204	205	210	182	207	207	194
7	206	212	181	190	204	196	201	204	200	222
8	196	203	200	206	180	209	207	215	195	191
9	210	204	184	194	220	177	209	218	190	189

SAMPLE OUTPUT III

Calculation of Success Ratios

with c = 337 and s = 8

Note: IIN = Success
IOUT = Failure

VALUES OF IIN AND IOUT

2	3	4	5	6	7	8	9	10	11	12	13	14	15
194	85	35	15	6	1	1	1	0	0	0	0	0	0
62	86	93	88	80	73	63	56	52	47	43	40	37	35
202	90	32	15	5	2	0	0	0	0	0	0	0	0
54	81	96	88	81	72	64	57	52	47	43	40	37	35
196	84	27	13	5	1	0	0	0	0	0	0	0	0
60	87	101	90	81	73	64	57	52	47	43	40	37	35
198	90	37	18	9	3	1	0	0	0	0	0	0	0
58	81	91	85	77	71	63	57	52	47	43	40	37	35
205	86	40	17	6	2	1	0	0	0	0	0	0	0
51	85	88	86	80	72	63	57	52	47	43	40	37	35
214	104	49	25	11	2	1	0	0	0	0	0	0	0
42	67	79	78	75	72	63	57	52	47	43	40	37	35
206	90	47	18	7	3	1	0	0	0	0	0	0	0
50	81	81	85	79	71	63	57	52	47	43	40	37	35
201	96	44	26	14	8	2	2	0	0	0	0	0	0
55	75	84	77	72	66	62	55	52	47	43	40	37	35
205	93	38	17	7	3	1	0	0	0	0	0	0	0
51	78	90	86	79	71	63	57	52	47	43	40	37	35
192	85	32	12	10	3	0	0	0	0	0	0	0	0
64	86	96	91	76	71	64	57	52	47	43	40	37	35
203	95	35	24	6	2	3	0	0	0	0	0	0	0
53	76	93	79	80	72	61	57	52	47	43	40	37	35
203	84	33	12	3	0	1	0	0	0	0	0	0	0
53	87	95	91	83	74	63	57	52	47	43	40	37	35
189	89	38	15	5	2	0	0	0	0	0	0	0	0
67	82	90	88	81	72	64	57	52	47	43	40	37	35
203	96	48	20	8	0	1	0	0	0	0	0	0	0
53	75	80	83	78	74	63	57	52	47	43	40	37	35
202	85	40	18	9	6	1	0	0	0	0	0	0	0
54	86	88	85	77	68	63	57	52	47	43	40	37	35
199	89	35	17	4	2	0	0	0	0	0	0	0	0
57	82	93	86	82	72	64	57	52	47	43	40	37	35
205	92	38	21	12	4	1	1	0	0	0	0	0	0
51	79	90	82	74	70	63	56	52	47	43	40	37	35
204	89	39	15	5	1	0	1	0	0	0	0	0	0
52	82	89	88	81	73	64	56	52	47	43	40	37	35
200	84	39	13	3	0	0	0	0	0	0	0	0	0
56	87	89	90	83	74	64	57	52	47	43	40	37	35
199	94	45	23	9	3	1	0	0	0	0	0	0	0

57	77	83	80	77	71	63	57	52	47	43	40	37	35
210	91	49	21	9	2	1	0	0	0	0	0	0	0
46	80	79	82	77	72	63	57	52	47	43	40	37	35
190	80	40	15	4	1	0	0	0	0	0	0	0	0
66	91	88	88	82	73	64	57	52	47	43	40	37	35
193	82	36	13	4	2	1	0	0	0	0	0	0	0
63	89	92	90	82	72	63	57	52	47	43	40	37	35
195	77	33	13	3	1	0	0	0	0	0	0	0	0
61	94	95	90	83	73	64	57	52	47	43	40	37	35
194	85	36	11	5	3	0	0	0	0	0	0	0	0
62	86	92	92	81	71	64	57	52	47	43	40	37	35
207	96	40	21	10	5	1	0	0	0	0	0	0	0
49	75	88	82	76	69	63	57	52	47	43	40	37	35
197	80	30	16	6	5	1	0	0	0	0	0	0	0
59	91	98	87	80	69	63	57	52	47	43	40	37	35
204	86	33	17	5	7	0	0	1	0	0	0	0	0
52	85	95	86	81	67	64	57	51	47	43	40	37	35
201	82	39	15	8	2	0	0	0	0	0	0	0	0
55	89	89	88	78	72	64	57	52	47	43	40	37	35
201	84	37	17	5	1	0	0	0	0	0	0	0	0
55	87	91	86	81	73	64	57	52	47	43	40	37	35
212	101	49	23	9	1	0	1	0	0	0	0	0	0
44	70	79	80	77	73	64	56	52	47	43	40	37	35
212	95	41	14	6	2	1	1	0	0	0	0	0	0
44	76	87	89	80	72	63	56	52	47	43	40	37	35

GROUP VALUES OF IIN AND IOUT

1	2	3
194	202	196
62	54	60
396	394	419
116	118	93
790	826	803
234	198	221
1616	1596	1596
432	452	452
85	90	84
86	81	87
175	174	190
167	168	152
349	376	357
335	308	327
725	716	689
643	652	679
35	32	27
93	96	101
67	64	89
189	192	167
131	180	138
381	332	374
311	299	319
713	725	705
15	15	13
88	88	90
30	31	42
176	175	164
61	86	65
351	326	347
147	135	134
677	689	690
6	5	5
80	81	81
11	14	17
161	158	155

25	38	26
319	306	318
63	52	49
625	636	639
1	2	1
73	72	73
3	4	4
145	144	144
7	15	8
289	281	288
22	18	14
570	574	578
1	0	0
63	64	64
1	1	2
127	127	126
2	5	5
254	251	251
7	7	4
505	505	508
1	0	0
56	57	57
1	0	0
113	114	114
1	2	0
227	226	228
3	0	2
453	456	454
0	0	0
52	52	52
0	0	0
104	104	104
0	0	0
208	208	208
0	0	0
416	416	416
0	0	0
47	47	47
0	0	0
94	94	94
0	0	0

82

188	188	188
0	0	0
376	376	376
0	0	0
43	43	43
0	0	0
86	86	86
0	0	0
172	172	172
0	0	0
344	344	344
0	0	0
40	40	40
0	0	0
80	80	80
0	0	0
160	160	160
0	0	0
320	320	320
0	0	0
37	37	37
0	0	0
74	74	74
0	0	0
148	148	148
0	0	0
296	296	296
0	0	0
35	35	35
0	0	0
70	70	70
0	0	0
140	140	140
0	0	0
280	280	280

THREE SUCCESS RATIOS FOR EACH GROUP FOR EACH N

1	2	3
0.757812	0.789062	0.765625
0.773437	0.769531	0.818359
0.771484	0.806641	0.784180
0.789062	0.779297	0.779297
0.497076	0.526316	0.491228
0.511696	0.508772	0.555556
0.510234	0.549708	0.521930
0.529971	0.523392	0.503655
0.273437	0.250000	0.210937
0.261719	0.250000	0.347656
0.255859	0.351562	0.269531
0.303711	0.291992	0.311523
0.145631	0.145631	0.126214
0.145631	0.150485	0.203883
0.148058	0.208738	0.157767
0.178398	0.163835	0.162621
0.069767	0.058140	0.058140
0.063953	0.081395	0.098837
0.072674	0.110465	0.075581
0.091570	0.075581	0.071221
0.013514	0.027027	0.013514
0.020270	0.027027	0.027027
0.023649	0.050676	0.027027
0.037162	0.030405	0.023649
0.015625	0.	0.
0.007812	0.007812	0.015625
0.007812	0.019531	0.019531
0.013672	0.013672	0.007812
0.017544	0.	0.
0.008772	0.	0.
0.004386	0.008772	0.
0.006579	0.	0.004386
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.
0.	0.	0.

FOOTNOTES

1. A serial correlation of lag $k=3$, for example, would indicate the correlation between the first and fourth, the second and fifth, the third and sixth numbers of the sequence, etc.

2. In the IBM Reference Manual on Random Number Generation and testing (11, p. 7) an auto-correlation coefficient is defined as

$$c_k \quad (1/N)\sum_{i=1}^{N} (x_i)(x_{i+k}).$$

Values of c_k are said to be approximately normally distributed with mean $= \frac{1}{4}$ and standard deviation $= .22/N^{\frac{1}{2}}$ for lags greater than zero and mean $= 1/3$ and standard deviation $= .30/N^{\frac{1}{2}}$ for a lag of zero. This method is not used here because support for the statistical statements could not be found.

3. In the classical definition of $R = \sum_{i=1}^{n} (x_i)(x_{i+k})$, where $i+k$ is greater than n, the formula becomes $R = \sum_{i=1}^{n} (x_i)(x_{i+k-n})$. This was originally established, however, for time series analysis in which limited data is available. It is here considered more realistic and a better indication of serial correlation in pseudo-random numbers to use the first formula for the calculation of R for any partial sampling from a generated sequence.

REFERENCES

1. Hull, T.E. and A.R. Dobell, "Random Number Generators", SIAM Review, 4(1962), pp. 230-251.

2. Hull, T.E. and A.R. Dobell, "Mixed Congruential Random Number Generators for Binary Machines", Journal of the Association for Computing Machinery, 11(1964), pp. 31-40.

3. Lehmer, D.H., "Mathematical Methods in Large Scale Computing Units", Proceedings of a Second Symposium (1949) on Large-Scale Digital Calculating Machinery, The Annals of the Computation Laboratory of Harvard University, 26(1951), pp. 141-146.

4. Peach, P., "Bias in Pseudo-Random Numbers", Journal of the American Statistical Association, 56(1961), pp. 610-618.

5. Rotenberg, A., "A New Pseudo-Random Number Generator", Journal of the Association for Computing Machinery, 7(1960), pp. 75-77.

6. Tocher, K.D., The Art of Simulation, D. Van Nostrand Company, Princeton (1963).

7. Coveyou, R.R., "Serial Correlation in the Generation of Pseudo-Random Numbers", Journal of the Association for Computing Machinery, 7 (1960), pp. 72-74.

8. Greenberger, Martin, "Notes on a New Pseudo-Random Number Generator", Journal of the Association for Computing Machinery, 8(1961), pp. 163-167.

9. Hoel, Paul G., Introduction to Mathematical Statistics, Wiley, New York (1962), pp. 341-345.

10. Wald, A. and Wolfowitz, "An Exact Test for Randomness in the Non-Parametric Case Based on Serial Correlation", Annals of Mathematical Statistics, 14(1943), pp. 378-388.

11. IBM Corporation, Random Number Generation and Testing, Reference Manual, C20-8011, New York (1959).

12. Marshall, A.W., "An Introductory Note", Symposium on Monte Carlo Methods, Wiley, New York (1956), pp. 1-14.

13. Davis, P. and P. Rabinowitz, "Some Monte Carlo Experiments in Computing Multiple Integrals", Mathematical Tables and Aids to Computation, 10(1956), pp. 1-8.

14. Cox, D.R., Planning of Experiments, Wiley, New York pp. 30-31.

15. Muir, James Wight, "The Effect of Scheduling on a Multi-Variate Stochastic System Evaluated Using Simulation Techniques", Master's Thesis, Lehigh University (1965).

VITA

Kenneth Joseph Lamport

Place of Birth:	New York City, New York
Date of Birth:	December 29, 1942
Parents:	Sol and Lydia Lamport
Education:	Lehigh University, Bethlehem, Pennsylvania
Degree:	Bachelor of Science in Industrial Engineering, June 1964